Breakout Discussion: Semantic-Aware Security

Moderators: XiaoFeng Wang (IU), William Enck (NCSU)

Scribe: Tudor Dumitras (UMD)

Why Semantics?

- Future of security technologies: Data-centric and intelligent

- In-depth understanding of data moves us toward this end
  - Detecting Online attack content aiming at human targets
  - Aggregating and analyzing attack-related information from different sources
  - Content analysis for supporting manual security analysis

Prior Attempts

- Permission check

e.g., Whyper: Towards automating risk assessment of mobile applications

- App behavior interpretation

e.g., Towards automatic generation of security-centric descriptions for android apps

- Compromised website detection

e.g, Seeking nonsense, looking for trouble: Efficient promotional-infection detection through semantic inconsistency search

- Cyber Intelligence gathering and analysis

e.g., Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence

e.g., FeatureSmith: Automatically engineering features for malware detection by mining the security literature

- Privacy policy compliance check

We are still not there, yet

- Existing research is ad-hoc, case by case
  - No systematic methodologies, no shared resources, no community effort to move the science forward

- Challenges for Semantic-Aware Security Research
  1. How to build the **Foundation** for the whole area?
  2. What are the application domains of these technologies?

Fundamental Questions?

- Can we directly apply existing NLP, Machine Learning and other techniques to support security research?

- What are the most important problems that need to be addressed in the area, so their answers can be utilized in different application domains?
  - Common techniques? Common datasets?

- What are the unique security challenges in the application of semantic technologies?
  - E.g., disinformation attack

New application domains

- Semantics-based risk detection and analysis
  - Cyber intelligence collection and analysis
  - Content-based detection
  - Description-functionality fidelity
- Support for attack response
  - Attack attribution
  - Risk communication
- Others?

How to foster the whole community

- Workshop to bring security and NLP/ML folks together?
- Annual community challenges to solicit the most promising technologies?
- Communication with the industry to identify the real world demands for semantics-aware security technologies?